

Individual Behavior under Evaluative Voting: A Comparison between Laboratory and *In Situ* Experiments

**Herrade Igersheim¹, Antoinette Baujard², Frédéric Gavrel³,
Jean-François Laslier⁴, Isabelle Lebon^{5, 6}**

March 26, 2015

(forthcoming in the book *Voting Experiments*, Springer)

Abstract. This paper compares two experimental methodologies for studying how individual voting behavior changes with respect to the choice of voting rule. We concentrate on different versions of Evaluative Voting. The results are based on two types of experimental protocol: a classical laboratory experiment with monetarily-induced preferences, and an *in situ* experiment run in parallel with the 2012 French presidential election. In the laboratory, individuals use the different rating scales in similar ways; but this is not the case *in situ*. The difference may be due to the different ways subjects interpret the proposed scales (in particular negative grades) when they concern real candidates. Finally, the paper discusses what each method can teach us about voter behavior. Notably, we highlight behavior that is in contradiction with the tenets of strict rationality. This phenomenon, which is very widely observed *in situ*, and where it may be explained by a motive to express oneself, is also observed in the lab, where it is still to be explained.

Classification JEL: D72, C93

Keywords: Voting, Field Experiment, In Situ Experiment, Laboratory Experiment, Evaluative Voting, Approval Voting, Strategic Behavior

¹ CNRS, BETA, Université de Strasbourg, 61, avenue de la Forêt Noire, 67085 Strasbourg Cedex, France; igersheim@unistra.fr.

² GATE L-SE, Université Jean Monnet, Université de Lyon.

³ CREM and Condorcet Center, Université de Caen Basse-Normandie.

⁴ CNRS, Paris School of Economics.

⁵ CREM and Condorcet Center, Université de Caen Basse-Normandie.

⁶ We warmly thank all those who contributed to our experiments. For the *in situ* experiment, see <http://www.gate.cnrs.fr/spip.php?article580#Merci>. For the lab experiment, we especially thank E. Priour and the LABEX in Rennes, and K. Boun My and the LEES in Strasbourg. This paper was first presented at the Montreal Voting Experiment Workshop held in March 2014; we are grateful to the project Making Electoral Democracy Work for their invitation and to the participants of the workshop for their comments. As regards funding, we thank the CREM, Chaire CNRS & UJM “Welfare economics” as well as the Foundation of the University of Strasbourg.

1. Introduction

This paper stems from a research project whose aim is to study individual behavior under evaluative voting. By “evaluative voting”, we refer to voting methods in which each voter grades each candidate on a pre-defined numerical scale, where the same grade may be given to several candidates, the sum of a candidate’s grades is his score, and the candidate who gets the highest score is elected. Different scales may be used: *approval voting*, for instance, in which the voter approves each candidate or not, is a particular case of evaluative voting with the simple scale (0,1). Scales such as (-1,0,1) or (0,1,2) may be used as well.

It is not clear how voters would use these scales, and we use the experimental method to gain better understanding on this point. There are different ways of doing experiments, and here we will study the insights we can draw from two different experimental contexts: *in situ*, and in the laboratory. *In situ*, people are invited to vote under alternative voting rules at the time and place of a real election, as if the alternative rule was the official rule: in particular, they are presented with the actual candidates for the election. In the laboratory, individuals vote according to controlled monetary preferences, and the “candidates” are just abstractions used to present the payment scheme. The present paper is thus a methodological contribution with respect to these two experimental methods.

Evaluative voting is studied in Normative Economics (Debreu 1960; d’Aspremont and Gevers 1977; Pivato 2013; Macé 2014) as the natural political counterpart of the utilitarian paradigm. Some authors have developed theoretical accounts regarding how individuals vote or should vote under different evaluative voting methods (Brams and Fishburn 1978; Hillinger 2004, 2005; Smaoui 2007; Gaertner and Xu 2012; Alcantud and Laruelle 2012; Núñez and Laslier 2014). But although the principle of additive evaluation is well grounded in theory and is widely used in practice (for example, in schools, sports, market research, feeling thermometers, etc.), the issue remains poorly studied from the empirical point of view.

The stream of literature devoted to *in situ* experiments since 2002 has contributed to filling this gap, at least in part (Grofman et al. 2011; Laslier 2011). It focused first on approval voting (Balinski et al. 2003; Laslier and Van der Straeten 2008; Alós-Ferrer and Granić 2012) and then turned to other evaluative rules (Baujard and Igersheim 2010; Baujard et al. 2011; Lebon et al. 2014).

An individual is likely to vote differently depending on the evaluative voting rule she faces (the set of allowed grades), and according to the context (laboratory or field). Here, we examine the observed modifications of individual behavior according to different evaluative voting rules when the context is changed. In this regard, our research problem can be seen as an analysis of two methodologies, *i.e.*, laboratory experiments, in the Experimental Economics tradition, and *in situ* experiments, on the basis of the comparison of different kinds of evaluative voting rules. A comparison between laboratory and field experimentation is provided by Jerit et al. (2013), specifically concerning media research. To the best of our knowledge, there is no such study about voting behavior.

This paper offers a comparative analysis of data from different experiments. Laboratory experiments were conducted in November 2008 in Rennes (Brittany, France) and in February 2014 in Strasbourg (Alsace, France). The protocol followed Van der Straeten et al. (2010), but with different voting rules. The *in situ* experiments took place during the French presidential election on April 22, 2012. Invitations to participate in the experiment were extended to the registered voters in five voting stations, and 2,340 voters eventually took part in the experiment. The rules under both tests were approval voting (henceforth, AV), and other variants of evaluative voting using the scales (0,1,2) and (-1,0,1) (henceforth, EV(0,1,2) and EV(-1,0,1)).

The paper is organized as follows. In the next section, we set out theoretical expectations leading to testable hypothesis. Section 3 briefly describes the experimental protocols used in the laboratory and *in situ*. Section 4 puts the theory of instrumental-strategic voting to the test in the two contexts. Section 5 compares how the different scales are used. We conclude in section 6 by offering some thoughts on what can be studied with these non-equivalent experimental methods.

2. Theoretical expectations

Experimental studies regarding voting rules first emerged in the mid-seventies (Fiorina and Plott 1978). Laboratory experiments on voting rest on the main principle of experimental economics, which consists in observing individual behaviors when individual preferences are monetarily induced and then comparing these data with theoretical predictions. Concretely, each participant is paid according to the candidate selected by the group with such or such voting method. In practice, participants are often students. The environment is defined by the experimentalists and depends on what they want to focus on. The precise reward scheme, the position of the candidates, the eventual interactions between the subjects, and the repetitions (or not) of the same voting rule, are parameters which enable the experimenter to reproduce different voting contexts and to model their specific characteristics. Two main questions are studied by laboratory experiments on voting rules: the comparison of aggregate results between voting rules, and the issue of strategic voting vs. sincere voting.

Even though all these works are highly pertinent in order to better understand individual behavior in voting contexts, their scope is limited by the artificiality of the voting environment typical of any laboratory experiment: sample bias, abstract candidates, monetary controlled preferences, etc. The recent stream of *in situ* research attempts to address precisely this issue, and as far as possible to observe “true” electoral preferences. Indeed, the *in situ* experiments, which are conducted in parallel with major official elections, aim at examining how real citizens (and not just students) behave at the time and place of national official elections: with real candidates, no repetition, etc. The goal of such experiments is to examine how voters react when confronted with alternative voting rules, and how they use them to effectively express their electoral preferences. The comparison of aggregate results and the issue of strategic vs. sincere voting are thus also addressed by this literature, and it is all the more relevant that real voters are expected to vote more sincerely in this context. In this paper, we focus on the issue of strategic vs. sincere voting: our theoretical predictions and the

subsequent empirical tests aim precisely at emphasizing the differences in individual behavior under the two contexts.

2.1 On instrumental-strategic voting

As stressed above, the main difference between laboratory and *in situ* experiments concerns voters' preferences. In the laboratory, preferences are monetarily induced and thus controlled through the experimental protocol. We expect individual behaviors to be totally driven by the aim of winning more money, which for these experiments means by the outcome of the election. This situation corresponds to a theoretical reference which gives a strictly instrumental and consequentialist value to the vote, with the "consequence" of a vote being the result of the election in a narrow sense: who is elected. We will refer to this rationality as "instrumental-strategic".

In situ, by contrast, individuals may be influenced by different kinds of elements likely to modify their strategic behavior, and to incite them to vote more sincerely. Indeed, voters may attribute an intrinsic value to the expression of their opinion rather than simply voting to influence the outcome of the experiment. They may consider the "outcome" of the election to concern not only who is elected but also the scores of the various candidates, and they may use their ballot as a device to influence these scores and send a political message. In addition, the fact that people know that they cannot influence the official election decreases the actual incentives at work under instrumental rationality. If this holds true for real elections, it is probably all the more so for *in situ* experiments. Cox (1997, p. 131) refers to this idea when he writes, "for example, some voters derive a consumption value from voting for their favored party in the first round, then such consumption values will overwhelm any instrumental values in large electorates (because the probability of one vote is infinitesimal). Non-instrumental voters of this kind will therefore not vote strategically." Notice, however, that expressive voters may well be strategic in the sense that deciding which signal to send may involve sophisticated reasoning and may well lead the voter to vote for a different candidate from her preferred one.

In any case, for the reasons developed above, we expect instrumental-strategic behavior to be more important in the laboratory than *in situ*. How to test this point?

Under evaluative voting, instrumental-strategic voting should, in theory, imply the following: a strategic voter should essentially focus on extreme grades, and thus there should be no major difference between the behavior of a strategic voter under evaluative rules and approval voting. The intuition behind this result is as follows: my vote makes a difference if it changes the election outcome – that is, in cases where there is a near-tie between several candidates and the number of points I am adding to their scores makes the difference. For each candidate (call her X), in order to decide how many points I should award to X, I consider the most serious race for the first place which might involve this candidate. This is a race between X and some other candidate Y. Then (if I am not indifferent) I should either try to make X win against Y, or lose against Y. The best thing to do is then to award to X either the maximal, or the minimal, number of points: Push or Pull, but not both! Núñez and Laslier (2014) have shown that, in the case of evaluative voting and for large elections, the

maximization of expected utility is essentially driven by the above reasoning, applied with respect to all candidates in turn following their expected scores.

Under approval voting, whether or not one approves a candidate for instrumental strategic reasons depends on whether or not one prefers her or him to the most likely winner of the election (Myerson and Weber 1993; Laslier 2009; Dellis 2010). In the case of AV, and according to this definition, there is no conflict between strategic and sincere voting (Van der Straeten et al. 2010). Therefore, in some of the analyses below, we will use AV as a benchmark to assess the quantitative importance of extreme grades as compared with intermediate grades under both EVs.

This theoretical point has a clear testable implication. Instrumental-strategic voters would not use intermediate grades. We thus expect the following:

H1-lab: In the laboratory, participants mostly use extreme grades.

We expect that voters are less instrumental-strategic in the field and more prone to use evaluative voting for expressive motives. Laboratory experiments are well suited for testing voters' strategic abilities, whereas, by comparison, *in situ* experiments cannot disentangle this ability from the willingness to be strategic. Even if some voters seize the opportunity to express their opinions in an extreme manner, it seems reasonable to expect that many of them will, on the contrary, use the offered grade scale in a way that shows that they have non-degenerate opinions on the candidates. We thus expect:

H1-field: Non-extreme grades are more often used *in situ* than in the laboratory.

We shall obtain that both hypotheses are confirmed by our data.

2.2 On using different scales

A candidate score under evaluative voting is computed by simple sum. As a consequence, the two grade scales (0,1,2) and (-1,0,1) would be equivalent if voters attach no intrinsic meaning to the labels “-1”, “0”, “1”, “2”. This is clear for purely instrumental-strategic voters, but should also be true, for instance, for voters interested in the relative scores of non-elected candidates. On the contrary, it may not be true for a voter who (i) attaches meaning to the grades themselves and (ii) uses her votes for the purpose of signaling.

In line with the received idea that the laboratory protocol is designed to isolate instrumental behavior, we thus expect the following:

H2-lab: The rules EV(0,1,2) and EV(-1,0,1) are equivalent in the laboratory.

We do not expect the equivalence to hold *in situ*, but here we can be more precise: *in situ*, voters were facing ten real candidates; most of them had no chance of winning, and some of them were probably not known by many voters. A voter who wishes to attach meaning to the numbers “-1”, “0”, etc. may wish to distinguish between candidates she does not know and candidates to whom she is opposed. We thus expect that the grade “0” in the scale (0,1,2) will be chosen for both unknown and rejected candidates and thus chosen more often than the grade “-1” in the scale (-1,0,1). We thus frame this hypothesis:

H2-field: Grade “0” under EV(0,1,2) is more frequent than grade “-1” under EV (-1,0,1).

Notice, however, that testing this hypothesis is not straightforward, because, given the protocol we implemented, the two scales were used in different locations. We will nevertheless manage to verify the hypothesis using approval voting (which was used in all locations) as a benchmark.

We will see that H2-lab cannot be refuted, and that H2-field is clearly accepted.

3. Experimental protocols

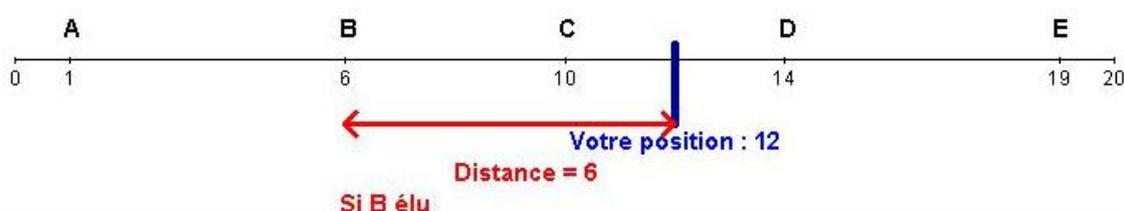
This section is devoted to the description of the experimental protocols in the two contexts: laboratory and *in situ*.

3.1. Protocol of the laboratory experiment

We conducted laboratory experiments in November 2008 in Rennes (Brittany, France) and in February 2014 in Strasbourg (Alsace, France). The protocol essentially replicates the one used in Van der Straeten et al. (2010) and can be described as follows:

Twenty-one subjects vote to elect “candidates”, *i.e.* alternatives located at five distinct points on a left–right axis that goes from 0 to 20. There are two extreme candidates (A, left and E, right), two moderate candidates (B, left and D, right) and one centrist (C). Each subject is assigned a position on the left–right axis, and knows that she will earn 20 euros minus the distance between her position and the elected candidate. In other words, the closer the position of the elected candidate is to the subject’s, the more money she earns. For instance, if the subject has the position 12 and if the elected candidate in position 6 wins, the distance between the subject and the candidate is 6. Then, for this election, the subject will earn $20 - 6 = 14\text{€}$ (see Figure 3.1). This reward scheme is kept identical throughout the experiment.

Figure 3.1. The position of the five candidates – Payoff for subject 12 if B wins



A series of five consecutive elections are conducted with the same voting method. At the end of each election, each subject knows the results (the scores of the candidates) immediately, as well as her own earnings according to her position. For each series, the subjects are randomly assigned a position on the left–right axis. There are 21 positions on this axis (from 0 to 20), and each subject has a different position.

Several series of five elections are held successively, with different voting rules. The subjects are never assigned the same position twice during the experiment. At the end of the session, one (in Rennes) or two (in Strasbourg) elections are randomly drawn and determine the payoffs (an average in the case of Strasbourg). At the beginning of the experiment, the

subjects are informed about the protocol as described above. The only thing they do not know is who holds which position for each series.

In Rennes, we performed six sessions and studied three voting rules: 2R (Two-round majority voting, the usual rule in France), AV, and EV(0,1,2). In Strasbourg, three sessions were performed and four voting rules were tested: 2R, AV and two kinds of evaluative voting, (EV(0,1,2) and EV(-1,0,1)).

Overall, AV and EV(0,1,2) were repeated 45 times (30 times in Rennes, 15 in Strasbourg), while EV(-1,0,1) was repeated 15 times. Table 3.1 presents the aggregate results per candidate. For instance, candidate C is elected 22.5 times out of 45 under AV, which represents 50.00% (the non-natural number of elections is easily explained: in case of ties between two candidates, $\frac{1}{2}$ point is given to each of them).

One may note that the extreme candidates A and E are never elected, but the differences observed between EV(0,1,2) and EV(-1,0,1) are not statistically different due to the small number of aggregate observations (45 for the former and 15 for the latter).⁷ Thus, no significant conclusion regarding H2-lab can be inferred from these data at the aggregate level. Everything that follows will be based on individual data observations.

Table 3.1. Aggregate results for the lab-experiment

	AV	EV(0,1,2)	EV(-1,0,1)
Centrist (C)	50.00%	66.67%	53.33%
Moderate (B or D)	50.00%	33.33%	46.67%
Extreme (A or E)	0.00%	0.00%	0.00%
Total	45 (100.00%)	45 (100.00%)	15 (100.00%)

3.2. Protocol of the In Situ experiment

The *in situ* experiment was conducted on April 22, 2012, during the first round of the 2012 presidential election in France. Five voting stations in three localities were selected to host the experiment: the two voting stations of the village of Louvigny, Normandy; one voting station in the city of Saint-Etienne, Rhône-Alpes; and two voting stations in the city of Strasbourg, Alsace. The voters were invited to test two voting rules, approval voting (AV) and one modality of evaluative voting: EV(0,1,2) in Saint-Etienne and EV(-1,0,1) in Louvigny.⁸

Two weeks before voting day, each registered voter received a letter by post, informing her of the experiment and of the principle of the two rules to be tested. On April 22, after (and conditional upon) their participation in the official vote, volunteers were asked to “vote experimentally”, i.e., to participate in the test of two alternative voting rules.

⁷ The centrist candidate is elected 30 times out of 45 (66.67%) on the one hand and 8 times out of 15 (53.33%) on the other. Even assuming that the 60 observations are independent, the difference $66.67 - 53.33$ is different from 0 with only 70% confidence.

⁸ Note that the modality tested in Strasbourg EV(0, ..., 20) is beyond the scope of this paper. See Baujard et al. 2014 or Lebon et al. 2014 for analyses in this regard.

Out of the approximately 5,000 registered voters, 2,340 volunteered to take part in the experiment. Note that this self-selected sample is not representative of the population of the voting places (there is a left-wing bias), and that the voting stations in the three cities also differ from each other.

4. Instrumental-strategic voting

As previously explained, our tool for the study of instrumental-strategic behavior will essentially be the use of extreme versus intermediate grades. We formulated two hypotheses. H1-lab: In the lab, participants use mostly extreme grades; and H1-field: Non-extreme grades are more often used in the field than in the lab. We discuss these hypotheses on the assumption that the participants of the experiments, in the lab as *in situ*, have taken the proposed voting rules seriously; this assumption is upheld by the observation that the proportion of inconsistent behavior is very low, both in the lab and *in situ*.

Statistics on the use of grades are provided in Table 4.1, which reads as follows: for the lab experiment under EV(-1,0,1) the grades -1, 0 and 1 represent respectively 52.32%, 10.10% and 37.59% of the total number of grades given to all the candidates. This table can be analyzed by taking the distribution of AV as a benchmark. Indeed, if the voters make use of approvals and grades in the very same way, they should grade the candidates under the three EVs as they did under AV, *i.e.*, for instance in the lab context 62.41% for the lowest grade and the rest for the highest grade, with no intermediate grade.

In the laboratory it is true that the intermediate grades are not often used: respectively 8.87%, and 10.10% for grade “1” under EV(0,1,2) and grade “0” under EV(-1,0,1). We must conclude that hypothesis H1-lab is validated, and we may recall that with three possible grades, there is a 90% concentration on the two extreme grades. But note that even in the laboratory individual behavior is not purely strategic.

Table 4.1. Use of approvals/grades (%)

		Lab-experiment	<i>In situ</i> experiment
AV	0	62.41%	73.49%
	1	37.59%	26.51%
		100.00%	100.00%
EV(0,1,2)	0	56.32%	59.55%
	1	8.87%	24.33%
	2	34.81%	16.12%
		100.00%	100.00%
EV(-1,0,1)	-1	52.32%	37.67%
	0	10.10%	39.72%
	1	37.59%	22.61%
		100.00%	100.00%

In situ the picture is very different. The share of intermediate grades is three or four times higher than in the lab (in Table 4.1, compare 8.87% with 24.33%, 10.10% with 39.72%). This corroborates hypothesis H1-field: Non-extreme grades are more often used *in situ* than in the laboratory.

A further look at the figures in Table 4.1 suggests more fine-grained conclusions, which are studied in next section.

5. The use of the different scales

When we compare EV(0,1,2) with EV(-1,0,1), the proposed scales differ with respect to the number of offered grades and with respect to the grades themselves. As explained in the theoretical section above, we are interested in comparing these two 3-grade scales in the laboratory. Looking again at Table 4.1, for the laboratory context one notices that the grades distributions for EV(0,1,2) and EV(-1,0,1) are very similar: compare 56.32%, 8.87%, 34.81% with 52.32%, 10.10%, 37.59%. Given that these figures come from different series of elections, yielding different histories, this similarity makes it impossible to refute hypothesis H2-lab: the scales (0,1,2) and (-1,0,1) are equivalent in the laboratory.

Turning to the field, the picture is again quite different. The statistical distributions of grades under EV(0,1,2) and under EV(-1,0,1) now differ. The lowest grade represents 59.55% and 37.67% respectively, in accordance with our hypothesis H2-field: grade “0” in the scale (0,1,2) is more frequent than grade “-1” in the scale (-1,0,1).

The problem here is that the two different scales were used by two different populations: EV(0,1,2) in Saint-Etienne and EV(-1,0,1) in Louvigny. The difference might thus be due to participants in Saint-Etienne being, for some reason, harsher on candidates than participants in Louvigny. Of course, we cannot totally rule out such an explanation, but we can control by reference to the opinion that each participant expressed on each candidate through approval voting.

Table 5.1 indicates how votes translate from AV to EV in the *in situ* experiments. It reads as follows: in Saint-Etienne, when a voter approves a candidate there is 3.41% chance that she gives this candidate grade “0”, a 39.98% chance that she gives grade “1”, and a 56.61% chance that she gives grade “2”. Hypothesis H2-field is then backed up by the observation that a candidate who is not approved receives grade “0” in the scale (0,1,2) with frequency 80.32% but receives grade “-1” in the scale (-1,0,1) with frequency only 50.27%.

The validation of H2-field calls for two slightly different interpretations. It can be interpreted as showing the specificity of negative grades. Indeed, for EV(-1,0,1), the grade “-1” is certainly considered by the participants as a “rejection” grade, and in a real context voters would reserve it only for the candidates they really dislike (Schwarz et al. 1991). But observe that a “rejection” grade is not necessarily negative, and in particular that the grade “0” might play this role too. In the last instance, it all depends on the way voters interpret the scale of grades of a voting rule.

Beyond this issue, the confirmation of H2-field proves indubitably that individual behavior differs depending on the voting rule in *in situ* contexts.

Table 5.1. Translation from AV to EVs in the *In Situ* experiments

Saint-Etienne: AV to EV(0,1,2)	Grade 0	Grade 1	Grade 2	366 ballots
From one approval under AV to...	3.41%	39.98%	56.61%	100.00%
From a “non-approval” under AV to...	80.32%	18.41%	1.28%	100.00%
Louvigny: AV to EV(-1,0,1)	Grade -1	Grade 0	Grade 1	896 ballots
From one approval under AV to...	1.88%	19.62%	78.50%	100.00%
From a “non-approval” under AV to...	50.27%	46.72%	3.01%	100.00%

6. Concluding remarks

The aim of this paper was to characterize the possible modifications of individual behavior according to different evaluative voting rules (AV, EV(0,1,2), EV(-1,0,1)) and in two distinct contexts: laboratory and *in situ*. At root, this research problem amounts to an evaluation of two methodologies devoted to experimental economics (laboratory versus field) on the basis of a comparison between evaluative voting rules employed in these two contexts. To do so, we conducted two experiments: on the one hand, a lab experiment in Rennes (November 2008) and Strasbourg (February 2014) where the voters’ preferences were monetarily induced; and on the other hand, an *in situ* experiment in parallel with the first round of the French presidential elections in April 2012. For the latter, the voters’ preferences depend on their real electoral preferences with respect to the candidates for this important national election.

We showed that the instrumental-strategic theory was a much better predictor of the observations in the laboratory than *in situ*. By contrast, *in situ*, voter behavior seems driven by their willingness to give their opinion on all the (real) candidates.

Going more into the details, we showed that in no context do individuals always act strictly as predicted by the theory of rationality: be it *in situ* or in the laboratory, the choice of the evaluation scale, although theoretically neutral, matters and influences individual behaviors.

Differences in individual behavior may come from the dichotomy between an intrinsic expressive value being attributed to the act of voting more often in the *in situ* context, as against the instrumental value attributed in the laboratory context. One might also consider whether they come from the different kinds of reasoning individuals use for the different evaluation scales.

Both methods are illuminating, and we can learn different things from them. By trying to control the participants’ objectives, the Experimental Economics methodology is designed to observe instrumental rationality at work. We thus learn from the lab how instrumental rationality works in elections. If instrumental rationality plays a major role in real elections,

this knowledge is important; if instrumental rationality is only marginal in real elections, this knowledge is interesting, but is only marginally useful.

We learned that voters in the lab use evaluations in a rather homogenous (across scales) and non-degenerate way: albeit not very often, they do use intermediate grades, which is in contradiction with strict rationality. This implies that intermediate grades are used even in the absence of the goal of self-expression. A cognitive mechanism which could explain this phenomenon is not known, and deserves to be studied. We might remark that this observation could have hardly been obtained in a real setting, where the use of intermediate grades would have been too easily attributed to expressive motives. This is an example of the value added by laboratory research.

By comparison, the *in situ* experiments showed differences in the use of the various evaluation scales, and in particular the specificity of negative grades, which could be interpreted as rejection grades. This was not seen in the laboratory, and can thus be attributed to expressive motives.

In order to deepen our understanding of these interesting and challenging conclusions, this work could be extended by more precisely characterizing the individual behaviors within each context, in order to bring to light the different kinds of reasoning individuals appeal to when faced with evaluative voting rules.

References

- Alcantud, José Carlos R., and Annick Laruelle. 2012. "To approve or not to approve, this is not the only question" *University of the Basque Country Working Paper Series*: IL, 63/12.
- Alós-Ferrer, Carlos, and Đura-Georg Granić, 2012. "Two field experiments on Approval Voting in Germany" *Social Choice and Welfare* 39: 171-205.
- Balinski, Michel, Rida Laraki, Jean-François Laslier, and Karine Van der Straeten. 2003. "Le vote par assentiment: une expérience" *Cahiers du Laboratoire d'Econométrie de l'Ecole Polytechnique*, n°2003-13.
- Baujard, Antoinette, Herrade Igersheim, Isabelle Lebon, Frédéric Gavrel and Jean-François Laslier. 2014. "Who's Favored by Evaluative Voting? An Experiment Conducted During the 2012 French Presidential Election" *Electoral Studies*, 34, 131-145.
- Baujard, Antoinette, and Herrade Igersheim. 2010. "Framed-field experiments on approval voting. Lessons from the 2002 and 2007 French Presidential elections". In *Handbook of Approval Voting*, eds. Jean-François Laslier and Remzi Sanver. Heidelberg: Springer, Chapter 15: 357–395.
- Baujard, Antoinette, Herrade Igersheim, and Thomas Senné. 2011. "An analysis of the political supply in the 2007 French Presidential election based on framed-field experimental data" *Annales d'Economie et Statistiques* 101-102, 149–186.
- Brams, Steven J., and Peter C. Fishburn. 1978. "Approval voting" *American Political Science Review*, 72, 831–847.
- Cox, Gary W. 1997. *Making Votes Count: Strategic coordination in the world's electoral systems*. Cambridge: Cambridge University Press.
- d'Aspremont, Claude and Louis. Gevers. 1977 "Equity and the informational basis of collective choice" *The Review of Economic Studies*, 44, 199–209.
- Debreu, G. (1960): "Topological methods in cardinal utility theory". In *Mathematical Methods in the Social Sciences*, Stanford University Press, 16–26.
- Dellis, Arnaud. 2010. "Policy moderation and endogeneous candidacy in approval voting elections". In *Handbook of Approval Voting*, Jean-François Laslier and Remzi Sanver (eds.). Heidelberg: Springer, pp. 431–453.
- Fiorina, Morris P. and Charles R. Plott (1978) "Committee decisions under majority rule: an experimental study" *American Political Science Review* 72:575–598.
- Gaertner, Wulf, and Yongsheng Xu. 2012. "A general scoring rule" *Mathematical Social Sciences*, 63, 193–196.
- Grofman, Bernard, Bernard Dolez, and Annie Laurent. 2011. *In Situ and Laboratory Experiments on Electoral Law Reform: French Presidential Elections*. Heidelberg: Springer.
- Hillinger, Claude. 2004. "On the possibility of democracy and rational collective choice" Discussion Paper, n° 2004-21, University of Munich.
- Hillinger, Claude. 2005. "The case for utilitarian voting" *Homo Oeconomicus*, 23, 295–321.

- Jerit, Jennifer, Jason Barabas, and Scott Clifford. 2013. "Comparing contemporaneous laboratory and field experiments on media effects" *Public Opinion Quarterly* 77: 256–282.
- Laslier, Jean-François. 2009. "The Leader Rule: a model of strategic approval voting in a large electorate" *Journal of Theoretical Politics* 21: 113–136.
- Laslier, Jean-François. 2011. "Lessons from *In Situ* tests during French elections." In *In Situ and Laboratory Experiments on Electoral Law Reform: French Presidential Elections*, Bernard Dolez and Bernard Grofman and Annie Laurent (eds.). Heidelberg: Springer, Chapter 5: 90–104.
- Laslier, Jean-François, and Karine van der Straeten. 2008. "Approval voting in the French 2002 presidential election: A live experiment" *Experimental Economics* 11: 97–195.
- Lebon, Isabelle, Antoinette Baujard, Frédéric Gavrel, Herrade Igersheim and Jean-François Laslier. 2014. "Individual behaviors facing different scales in evaluation voting", working paper.
- Macé, Antonin. 2014. "Voting with evaluations: when should we sum, what should we sum?" Working paper, Ecole Polytechnique.
- Myerson Roger B. and Robert J. Weber. 1993. "A theory of voting equilibria" *American Political Science Review* 87: 102–114.
- Núñez, Matias, and Jean-François Laslier. 2014. "Preference intensity representation: Strategic overstating in large elections" *Social Choice and Welfare* 42: 313–340.
- Pivato, Marcus. 2013. "Formal utilitarianism and range voting," *Mathematical Social Sciences* 67: 50–56.
- Schwarz, Norbert, Bärbel Knäuper, Hans-J. Hippler, Elisabeth Noelle-Neumann and Leslie Clark. 1991. "Numeric values may change the meaning of scale labels", *The Public Opinion Quarterly* 55: 570-582.
- Smaoui, Hatem. 2007. "Le système de vote par note à trois niveaux : étude axiomatique." Working Paper BETA-CREM.
- Van der Straeten, Karine, Jean-François Laslier, Nicolas Sauger, and André Blais. 2010. "Strategic, sincere, and heuristic voting under four election rules: an experimental study" *Social Choice and Welfare* 35: 435–472.